Appl. Statist. (1989)
38, *No.* 1, *pp.* 139–147

Rotation of Ill-defined Principal Components

By I. T. JOLLIFFE†

University of Kent, Canterbury, UK

[Received May 1987. Revised April 1988]

SUMMARY

Following a principal component analysis, it is a fairly common practice to rotate the first few components. The idea is that the rotated versions of the components, while jointly conveying the same information as the unrotated components, may be much easier to interpret. There are some drawbacks to rotation, but it is argued in this paper that rotation can be a useful procedure. However, rotation is best reserved for groups of principal components (not necessarily the first few) whose variances are nearly equal.

Keywords: Factor analysis; Ill-defined principal components; Oblique rotation; Principal component analysis; Rotation; Simple structure; Varimax rotation

1. Introduction

The most common use of principal component analysis is to reduce the dimensionality of a data set by replacing the p variables which have been measured by a much smaller number m of principal components. The value m is chosen so that most of the variation in the p original variables is accounted for by the first m components. It is useful to be able to interpret the m retained components; where this is difficult, rotation of the first m components in an attempt to achieve 'simple structure' may improve matters. In some fields of application rotation is almost routine, so much so that some standard statistical computer packages (e.g. BMDP and SPSS-X) will automatically rotate principal components unless a 'no-rotation' option is chosen. Rotation of principal components in this way is often thought of as a form of factor analysis, although to do so is misleading—see Jolliffe (1986), ch. 7. Rotating the first m principal components can improve their interpretability, but it suffers from several drawbacks. First, there are several essentially arbitrary choices to be made in implementing the rotation, and changing any of these choices may change the results to a lesser or greater extent. These choices include the following.

(a) Many different criteria have been suggested for rotation. For example, Cattell (1978) and Richman (1986) give non-exhaustive lists of 11 and 19 rotation criteria respectively.

†*Address for correspondence*: Mathematical Institute, University of Kent, Cornwallis Building, Canterbury, Kent, CT2 7NF, UK.

 0035–9254/89/38139 $2.00

(b) The kth principal component z_k is a linear function $\boldsymbol{\alpha}_k'\mathbf{x}$ of the vector $\mathbf{x}$ of the original p variables. (For simplicity, our notation will not distinguish between sample and population principal components. It should be clear from the context which is meant.) The elements of $\boldsymbol{\alpha}_k$ may all be multiplied by an arbitrary constant, without changing the nature of the kth component. Thus, a normalization constraint must be imposed on $\boldsymbol{\alpha}_k$ to specify z_k uniquely. Different normalization constraints may lead to different, and non-equivalent, rotated components. There are many different possible constraints, but two are more often adopted than others. The first is $\boldsymbol{\alpha}_k'\boldsymbol{\alpha}_k = 1$, which arises naturally in the derivation of principal components and has the property of preserving distances. The alternative $\boldsymbol{\alpha}_k'\boldsymbol{\alpha}_k = \lambda_k$ is adopted in many computer packages because elements of $\boldsymbol{\alpha}_k$ can be interpreted as correlations between the kth component and the original variables.

(c) Increasing the value of m from m_1 to m_2, say, simply adds $m_2 - m_1$ extra unrotated components, leaving the original m_1 components unchanged. However, after rotation, none of the m_2 rotated components need bear any resemblance to the m_1 rotated components.

A fourth drawback of rotation is the loss of a defining property of principal components.

(d) The principal components successively have maximum variance, and if the first m components are rotated, then the m rotated components jointly account for the same (maximal) amount of variation as the m unrotated components. However, the m rotated components no longer individually represent maximal amounts of variation. The variation accounted for by the first m components is typically spread much more evenly between individual components after rotation than before. Thus we may lose information on the most dominant individual sources of variation in the data.

Instead of rotating the first m components, the idea of rotating all p principal components has been found useful in regression (Hawkins, 1973; Jeffers, 1981) and in outlier detection (Hawkins and Fatti, 1984). The drawbacks listed can still cause problems in this situation, except for (c), which is no longer relevant.

There is a context, however, in which rotation can be used to simplify structure, without the drawbacks which are usually encountered. This occurs when we rotate subsets of components (not necessarily the first few) whose variances are nearly equal. The next section discusses this idea further, and examples of its use are given in Section 3.

2. Rotation of Ill-defined Components

Consider a set of q successive principal components whose variances (eigenvalues) $\lambda_{s+1}, \lambda_{s+2}, \ldots, \lambda_{s+q}$ are nearly equal. Then it is well known that the vectors $\boldsymbol{\alpha}_{s+1}$, $\boldsymbol{\alpha}_{s+2}, \ldots, \boldsymbol{\alpha}_{s+q}$ of coefficients for individual components $z_{s+1}, z_{s+2}, \ldots, z_{s+q}$ in this set are unstable. For example, Krzanowski (1984) implicitly shows that small changes in any of the λ_k values in this subset can lead to large changes in the corresponding $\boldsymbol{\alpha}_k$ values. Furthermore, deletion of single observations from the data set can also cause large changes in the $\boldsymbol{\alpha}_k$ values in such a subset—see Pack *et al.* (1988). However, the subspace defined by the q components is stable, provided that λ_s and λ_{s+1} are well separated, as are λ_{s+q} and λ_{s+q+1}.

Thus, we can divide the p principal components derived from a set of p variables into subsets such that variances of components within subsets are nearly equal, whereas variances in different subsets are well separated. Then each subspace defined by one of our subsets of components is relatively stable or well defined, whereas individual components within a subset are unstable or ill defined.

Given a particular well-defined subspace, it will be an advantage to be able to interpret this subspace as simply as possible. Rotation of components within a subspace to obtain simple structure is one way of achieving this. Furthermore, because of the closeness of the eigenvalues corresponding to the rotated components, some of the drawbacks of rotation listed in Section 1 no longer have the same force that they do when eigenvalues are well separated.

In particular, problem (d) is no longer a worry because within our subsets the variance is already fairly equally spread between components before rotation. Any component which represents a dominant one-dimensional source of variation will have a greatly different variance from adjacent components, and so will not be considered for rotation.

Drawback (b) no longer has much relevance either; the usual normalization constraints relate $\boldsymbol{\alpha}_k$ either to a constant or to a function of λ_k. If all λ_ks in a rotated subset are similar, then the specific choice of a normalization constraint has very little effect on the relative values of the elements of the $\boldsymbol{\alpha}_k$s within the subset.

Regarding restriction (c), we still have to decide on how to divide our p components into subsets for rotation, but this should often be a less arbitrary choice than deciding how many components to retain to account for most of the variation in a data set. Many suggestions have been put forward for deciding how many components to retain, but none yet seems to offer much advantage over simple *ad hoc* rules—see Jolliffe (1986), ch. 6.

Finally, drawback (a) remains. The choice of which rotation criterion to use is still arbitrary but should reflect how we define our objective of simple structure among the components. Further discussion of rotation criteria is given later in the paper.

3. Examples

3.1. *Artistic Qualities*

The small artistic qualities data set was discussed by Davenport and Studdert-Kennedy (1972). Four subjective measurements of the artistic qualities 'composition', 'drawing', 'colour' and 'expression' are available for 54 painters. In the original paper, principal components were found using the covariance matrix. Results presented here are for the correlation matrix, but, because of the similar-sized variances in this data set, there is little difference between the analyses. In general, there could be substantial differences between results for covariance and correlation matrices. The examples in this paper are all for correlation matrices, because these are more commonly used in practice, but there is nothing to prevent similar methodology being used for covariance matrices.

Returning to the present example, the analysis based on the correlation matrix for these data, the four eigenvalues were 2.27, 1.04, 0.40 and 0.29. In a conventional analysis it would be noted that the two dominant sources of variation given by the first

TABLE 1
Artistic qualities data: coefficients of components 3 and 4

	Unrotated		Rotated	
	Component 3	Component 4	Component 3	Component 4
Composition	−0.59	−0.41	−0.27	−0.66
Drawing	0.60	−0.50	0.78	−0.09
Colour	0.49	−0.22	0.53	0.09
Expression	0.23	0.73	−0.21	0.74
Percentage of total variation	10.0	7.3	9.3	8.0

two components together account for 83% of the total variation. These two components would therefore be retained and perhaps rotated. However, the eigenvalues associated with these two components are well separated from other eigenvalues and the components are therefore well defined. By contrast, eigenvalues 3 and 4 are much closer to each other, so that components 3 and 4 are less well defined. In Table 1, we present the coefficients of these two components, both before rotation and after rotation using the varimax criterion, as implemented by BMDP. Also given in Table 1 are the percentages of total variation accounted for by each component, both before and after rotation. The choice of the varimax criterion is arbitrary and for illustration only. Other criteria have been tried on some of the examples discussed in this paper. Table 2 compares three rotation methods: varimax and quartimax, which are orthogonal, and quartimin, which is oblique. The differences in this table are trivial and confirm the empirical finding in factor analysis that the choice of rotation method often has less effect than other choices, such as the number of factors to rotate—see, for example, Stewart (1981), who cites three references from the behavioural sciences literature. However, it is possible for different rotation criteria to produce different results, especially when oblique, as well as orthogonal, rotations are considered. Ideally, careful thought is necessary when choosing a rotation criterion, to make sure that it truly reflects the user's idea of what simple structure means.

The principal components which are rotated in Table 1 have the normalization criterion $\boldsymbol{\alpha}_k'\boldsymbol{\alpha}_k = 1$. If the alternative criterion $\boldsymbol{\alpha}_k'\boldsymbol{\alpha}_k = \lambda_k$ is used, there is little difference in the results. The angles between the vectors of the coefficients for the two normalization constraints are 6° and 2° for the third and fourth rotated components respectively.

TABLE 2
Artistic qualities data: different rotation criteria for components 3 and 4

Varimax		Quartimax		Quartimin	
Component 3	Component 4	Component 3	Component 4	Component 3	Component 4
−0.27	−0.66	−0.27	−0.66	−0.25	−0.66
0.78	−0.09	0.78	−0.10	0.78	−0.11
0.53	0.09	0.53	0.09	0.53	0.08
−0.21	0.74	−0.20	0.74	−0.22	0.74

It is clear from Table 1 that rotation of components 3 and 4 has considerably simplified their joint structure. In particular, the rotated component 4 is almost a pure difference between expression and composition, whereas there is no such simple interpretation for either of the unrotated components. Rotated component 3 is largely the sum of colour and drawing, and again has a simpler structure than either of the unrotated components.

At this stage it may be asked whether these components are of any interest, since most of the variation is accounted for by components 1 and 2. The last few, low variance, components define linear functions of $\mathbf{x}$ with the smallest possible variances, so that they are useful in identifying whether there are any near-constant relationships between the p variables in $\mathbf{x}$. There are many ways in which low variance, or minor, components can be used (see, for example, Hawkins and Fatti (1984)), and if these components have simple interpretations their usefulness will be enhanced. Rotation of minor components is therefore a worthwhile exercise.

3.2. *Facial Spots*

The facial spots data refer to the numbers of facial spots in five categories of severity (i.e. five variables) for 34 individuals. In a principal component analysis based on the correlation matrix for these data, the eigenvalues are 2.70, 0.87, 0.67, 0.46 and 0.31. None of these are particularly close to one another, so rotation is not strongly suggested. However, there is a clear outlier in the data set, and when this individual is removed the picture changes. The outlier violates the correlation structure of the remainder of the data; all variables are positively correlated, whereas the outlying individual has very low values on four of the five variables, but one of the highest values on variable 5. Interestingly, in the light of what follows, this outlier shows up very clearly on the second principal component but is not the most extreme observation on any of the other components or on any of the original variables. The second component is dominated by variable 5 but also contrasts it with some of the other variables—see Table 3.

If the outlier is omitted and the principal component analysis is repeated for the new correlation matrix, the eigenvalues are now 2.79, 0.74, 0.70, 0.46 and 0.31. In addition, the nature of the second and third principal components apparently changes when the outlying individual is omitted, whereas the first, fourth and fifth components are relatively unchanged. To quantify this statement, the coefficients of the second and third components with and without the outlying individual are given in Table 3.

TABLE 3
Facial spots data: coefficients of components 2 and 3

Unrotated				*Rotated*	
Outlier included		*Outlier omitted*		*Outlier omitted*	
Component 2	*Component 3*	*Component 2*	*Component 3*	*Component 2*	*Component 3*
−0.23	0.69	0.53	−0.49	0.69	−0.20
−0.22	0.22	−0.03	−0.49	0.19	−0.46
0.01	−0.56	−0.51	0.13	−0.52	−0.12
−0.12	−0.36	−0.36	0.22	−0.42	0.03
0.94	0.18	0.57	0.68	0.20	0.86

The second and third components have their order switched, but there are also differences in structure. The angles between the vectors of coefficients defining the two pairs of components with and without the outlier are 34° and 29°, whereas the angles between the pairs of vectors corresponding to components 1, 4 and 5 with and without the outlier are only 6°, 17° and 13° respectively. It appears, therefore, that omitting the outlying observation has a substantial effect on the structure of components 2 and 3. However, the eigenvalues for these two components with the outlier omitted are very close to each other, so that the individual eigenvectors are not well defined. Rotation of components 2 and 3 without the outlier, using varimax rotation and the normalization constraint $\boldsymbol{\alpha}_i'\boldsymbol{\alpha}_i = 1$ gives rotated coefficients as in the final two columns of Table 3. It can be seen that, apart from an (arbitrary) change of order, the rotated components have become more similar to components 2 and 3 using all the data. The angles between the corresponding pairs of vectors are now reduced to 18° and 5°.

To see what is happening in this example, recall that the outlying observation is extreme with respect to the second component. However, it is not sufficiently extreme to be mainly responsible for the direction of this component; without the observation the component is still important, though its variation is reduced to approximately that of the third component. Thus, the two-dimensional plot with respect to components 2 and 3 remains virtually unchanged apart from removal of the outlier, but the variation is now almost circular (i.e. all directions in the two-dimensional space defined by components 2 and 3 have very nearly equal variance) leading to arbitrariness in the definition of individual components within this subspace.

Removal of the outlier has therefore opened up the possibility of an arbitrary rotation, making the outlier seem very influential for components 2 and 3, when actually its influence is quite small. Varimax rotation has revealed that the original structure is still present, although this revelation was fortuitous and would not have happened if the original structure had not been simple. To look specifically for the original structure, we could adapt the procedure so that, instead of using varimax rotation, we rotate as closely as possible to the original vectors of coefficients.

3.3. *English Counties*

As a final illustration, we consider an example with rather more variables. The 12 variables are demographic characteristics of 46 English counties; the data were first analysed by E. Stone in an unpublished undergraduate dissertation, and some of the results were reported in Jolliffe (1986), pp. 167–170. Using the correlation matrix, the eigenvalues are

6.27, 2.53, 1.16, 0.96, 0.41, 0.26, 0.21, 0.12, 0.04, 0.03, 0.02 and 0.01.

Possible candidate subsets of components for rotation include components 3 and 4 (eigenvalues 1.16 and 0.96), components 6 and 7 (eigenvalues 0.26 and 0.21) and components 9–12 (eigenvalues 0.04–0.01). We shall see that these three subsets each exhibit slightly different behaviour when rotation using varimax is attempted.

First, components 3 and 4 already have a relatively simple structure, and rotation makes very little difference; the angle between the unrotated and rotated vectors of principal component coefficients is less than 8°. Turning to the last four components, coefficients before and after rotation using varimax and the normalization constraint $\boldsymbol{\alpha}_i'\boldsymbol{\alpha}_i = 1$ are given in Table 4.

TABLE 4
English counties data: coefficients of components 9–12

Before rotation				*After rotation*			
9	10	11	12	9	10	11	12
0.05	0.71	0.20	−0.21	0.03	0.75	0.08	0.15
0.62	0.01	0.16	−0.20	−0.05	0.09	0.67	−0.01
0.64	−0.10	0.23	−0.09	0.07	−0.04	0.69	0.01
0.13	−0.03	0.33	0.60	0.68	−0.14	0.09	0.03
−0.03	0.21	0.26	0.65	0.72	0.07	−0.12	−0.02
−0.15	0.02	−0.06	0.04	−0.00	0.00	−0.17	−0.01
−0.01	−0.48	0.17	0.05	0.03	−0.47	0.08	0.19
−0.01	−0.03	−0.04	−0.07	−0.09	−0.02	−0.01	−0.00
−0.10	−0.13	−0.02	−0.05	−0.09	−0.12	−0.07	0.05
−0.26	−0.32	0.53	−0.23	−0.06	−0.24	−0.01	0.66
0.08	0.11	−0.41	0.15	−0.02	0.06	−0.08	−0.44
0.28	−0.26	−0.48	0.21	−0.04	−0.32	0.11	−0.55

Apart from an arbitrary change of order, there is a clear correspondence between the components before and after rotation. Components 9, 10, 11 and 12 before rotation correspond respectively to components 11, 10, 12 and 9 after rotation. However, the correspondence is by no means exact; the angles between corresponding pairs of vectors are 21°, 12°, 31° and 31°. These large angles reflect the considerable simplification that has been achieved, especially for the last two components, while retaining the main features of the original components.

Finally in this example, consider components 6 and 7—see Table 5. Rotation changes the character of these components quite noticeably, although the angle of rotation (31°) is no bigger than some of those for components 9–12. In this case, both of the original components have many non-trivial coefficients; rotation produces one component which is much simpler, but the second rotated component is still difficult to interpret, with seven or eight non-trivial coefficients.

TABLE 5
English counties data: coefficients of components 6 and 7

Before rotation		*After rotation*	
Component 6	*Component 7*	*Component 6*	*Component 7*
0.17	0.32	0.36	0.02
−0.35	0.13	−0.07	0.36
0.17	−0.09	0.02	−0.19
0.11	−0.08	−0.02	−0.14
−0.08	0.11	0.05	0.13
−0.30	0.64	0.39	0.58
0.21	0.24	0.32	−0.06
0.71	0.00	0.37	−0.60
0.15	0.25	0.29	−0.00
−0.30	−0.08	−0.22	0.21
−0.24	−0.41	−0.47	−0.00
−0.00	0.40	0.34	0.21

4. Discussion

The examples in the previous section have shown that rotation of ill-defined principal components can provide simplification and therefore easier interpretation, without some of the drawbacks which are associated with the more usual rotation of the first few components. The second example also showed that rotation can provide additional insight into the influence of individual (outlying) observations. However, there are several questions regarding the technique which have yet to be investigated.

It was noted in Section 2 that two of the four drawbacks of rotation listed there essentially disappear if the eigenvalues of the rotated components are close together, but two drawbacks involving arbitrary choices remain and would benefit from further investigation. First, there is the question of which rotation criterion to use. In the examples given, varimax was arbitrarily chosen, but other criteria, using different definitions of 'simplicity' could equally well have been selected. In particular, it is not necessary to restrict the choice to orthogonal rotation. For q exactly equal eigenvalues, the corresponding q-dimensional subspace can be defined by any q vectors, not necessarily orthogonal, which form a basis for the subspace. Thus, oblique rotation is a possibility and may simplify coefficients of rotated components more than is possible using orthogonal rotation. Alternatively, in some cases, as in the second example given, we may be interested in rotating, not to achieve simple structure, but to try to approach as closely as possible to a specified set of components —Krzanowski's (1979) work on between-groups comparison of principal components is relevant here.

The other arbitrary choice to be made is which subsets of principal components to rotate, i.e. when are eigenvalues sufficiently close to make the corresponding components ill defined. An *ad hoc* test based on asymptotic distributional results for individual eigenvalues and eigenvectors has been used extensively in the climatological literature since it was suggested by North *et al.* (1982). The rule is that two consecutive eigenvalues should have a separation greater than the sum of estimates of their respective standard errors; these estimates are $\lambda_k(2/n)^{1/2}$ where n is the sample size. More formal tests are available for testing equality of subsets of eigenvalues, again based on asymptotic theory—see, for example, Davis (1977). Expressions for the sensitivity of eigenvectors to changes in the eigenvalues (Krzanowski, 1984) and to deletion of individual observations (Critchley, 1985; Pack *et al.*, 1988) are both functions of $(\lambda_k - \lambda_{k+1})^{-1}$. This suggests that a simple rule for determining whether λ_k and λ_{k+1} are sufficiently close should perhaps be based on $\lambda_k - \lambda_{k+1}$, rather than on anything more complicated, such as $(\lambda_k - \lambda_{k+1})/\lambda_k$, which is roughly the basis of the rule of North *et al.* (1982). A referee has suggested that cross-validation might be used to obtain confidence intervals for $\lambda_k - \lambda_{k+1}$, leading to the conclusion that λ_k and λ_{k+1} are well separated if the interval does not contain zero. However, the relative usefulness in practice of these various suggestions has yet to be assessed.

Acknowledgements

Helpful comments from the referees have improved the discussion in several places in the paper.

References

Cattell, R. B. (1978) *The Scientific Use of Factor Analysis in Behavioural and Life Sciences*, p. 136. New York: Plenum.

Critchley, F. (1985) Influence in principal components analysis. *Biometrika*, **72**, 627–636.

Davenport, M. and Studdert-Kennedy, G. (1972) The statistical analysis of aesthetic judgment: an exploration. *Appl. Statist.*, **21**, 324–333.

Davis, A. W. (1977) Asymptotic theory for principal component analysis: non-normal case. *Austral. J. Statist.*, **19**, 206–212.

Hawkins, D. M. (1973) On the investigation of alternative regressions by principal component analysis. *Appl. Statist.*, **22**, 275–286.

Hawkins, D. M. and Fatti, L. P. (1984) Exploring multivariate data using the minor principal components. *Statistician*, **33**, 325–338.

Jeffers, J. N. R. (1981) Investigation of alternative regressions: some practical examples. *Statistician*, **30**, 79–88.

Jolliffe, I. T. (1986) *Principal Component Analysis*. New York: Springer.

Krzanowski, W. J. (1979) Between-groups comparison of principal components. *J. Amer. Statist. Ass.*, **74**, 703–707; correction **76**, 1022.

——(1984) Sensitivity of principal components. *J. R. Statist. Soc.* B, **46**, 558–563.

North, G. R., Bell, T. L., Cahalan, R. F. and Moeng, F. J. (1982) Sampling errors in the estimation of empirical orthogonal functions. *Month. Weath. Rev.*, **110**, 699–706.

Pack, P., Jolliffe, I. T. and Morgan, B. J. T. (1988) Influential observations in principal component analysis: a case study. *J. Appl. Statist.*, **15**, 37–50.

Richman, M. B. (1986) Rotation of principal components. *J. Climatol.*, **6**, 293–335.

Stewart, D. W. (1981) The application and misapplication of factor analysis in marketing research. *J. Mktng Res.*, **18**, 51–62.